Available online at www.sciencedirect.com

ScienceDirect

IFAC Papers Online CONFERENCE PAPER ARCHIVE

IFAC-PapersOnLine 48-28 (2015) 1232–1237

A Time-Domain Fractional Approach for Wiener-Hammerstein Systems Identification

G. Giordano * **J. Sjöberg** *

* *Chalmers University of Technology, Department of Signals & Systems, SE412 96 Gothenburg, Sweden (e-mail: Giuseppe.Giordano@chalmers.se, Jonas.Sjoberg@chalmers.se)*

Abstract: This paper describes a new approach to initialize Wiener-Hammerstein models for iterative prediction error minimization. The key idea is to parameterize the division of poles and zeros of the best linear approximation (BLA), between the two linear subsystems. Taylor expansion is used to handle the parameterization in the time-domain. Results regarding uniqueness of the initial estimate are proved on a low order example. The initial estimate generated with this approach allows to avoid problems with local minima and it is used for the final optimization in all model parameters.

Keywords: Nonlinear System Identification, Block-oriented Nonlinear Models, Wiener-Hammerstein Model, Initial Estimates

Fig. 1. A Wiener-Hammerstein model structure.

1. INTRODUCTION

Nonlinear systems are present everywhere in real life and, though linear system theory and linear system identification methods have been often successfully applied in this field, the need of nonlinear system identification methods is increasing. One approach to nonlinear system identification is to work with block-oriented nonlinear models, defined as the interconnection of two kind of elements: linear, time-invariant dynamic and static nonlinear blocks. One example is the Wiener-Hammerstein (W-H) model, consisting of two linear dynamic blocks and a static nonlinearity in the middle, see Fig. 1. The identification of the W-H systems is challenging due to the presence of two dynamic systems whose contributions to the system behaviour are not easily separable. Generally, in system identification, the prediction error estimate gives an asymptotic efficient estimator when the number of data goes to infinity, see Ljung (1999) and Söderström and Stoica (1988). However, in practical cases, the cost function can have many local minima and, therefore, the main challenge is to generate a good initial estimate for the identification procedure, in order to increase the chances that the estimate converges to the global minimum. Many approaches deal with the minimization of the prediction error and the initial estimate problem is solved in different ways.

In Wills and Ninness (2012) a random, stable, initialization of G_W and G_H is performed. The authors show that there are many local minima, so the estimation needs to be repeated several times with different starting values, in order to decrease the risk of termination in local minima.

In the case of W-H model, it is relatively easy to estimate the product of transfer functions $G_W\,G_H$ through the Best Linear Approximation (BLA). Thus, other approaches are built on the splitting of the BLA in the two linear subsystems. Usually, in this case, the drawback is that a large number of optimization problems are needed.

In Sjöberg and Schoukens (2012), the "brute force" method performs an iterative optimization for each of the possibly pole/zero partition (combinatorially growing in the model order). This requires high computational time if the model order is greater than 10.

In Sjöberg et al. (2012), a least squares approach is used where the input signal is filtered through the basis functions of the first linear block, containing the poles of the BLA, and the output is filtered through the basis functions describing the inverse of the BLA, based on the zeros of the BLA. This approach does not have the drawback with exponential increasing of the computational time with respect to the model order but the linear blocks will have higher order than necessary.

More recently, in Vanbeylen (2014), the fractional approach is introduced for initialization. All poles and zeros of the BLA are placed both in G_W and G_H, and initialized using fractional exponents. The estimation problem is solved in frequency domain, where the fractional exponents can be easily handled. The method provides good results but the estimation problem is nonlinear in the fractional exponents. It is solved with an interative minimization algorithm which can be caught in local minima.

The contribution of this work is a time-domain formulation of the fractional approach. In time-domain it is possible to rewrite the predictor of the system as a linear regres-

10.1016/j.ifacol.2015.12.300

sion where the coefficients are algebraic expressions in the fractional exponents and in the parameters of the static nonlinearity. Then, given these coefficients, the fractional exponents can be analytically computed. On a low order example, uniqueness of the solution in the fractional exponents is shown.

The paper is organized as follows. A description of the model setting and the model structure is given in Section 2. In Section 3, the fractional approach is presented in details together with the uniqueness results. Section 4 summarizes the main steps of the new identification procedure and, in Section 5, a simulation example is used to test the new method.

2. PROBLEM FORUMLATION

2.1 Model Structure

The classical model structure of W-H type in the discrete-time is described by

$$v(t) = G_W(q^{-1}, \theta_W)u(t), \tag{1}$$

$$w(t) = f(\theta_{NL}, v(t)), \tag{2}$$

$$\hat{y}(t) = G_H(q^{-1}, \theta_H)w(t), \tag{3}$$

where $\hat{y}(t)$ is the predicted output of the system, and $G_W(q^{-1}, \theta_W)$ and $G_H(q^{-1}, \theta_H)$ are linear time invariant transfer functions in the delay operator q^{-1}. The function f is a static nonlinearity parameterized with θ_{NL}. All the parameters in the model can be collected in the vector

$$\theta = [\theta_W, \theta_{NL}, \theta_H]. \tag{4}$$

The two linear parts of the model can be further described by

$$G_W(q^{-1}, \theta_W) = \frac{\sum_{k=0}^{n_B^W} b_k^W q^{-k}}{1 + \sum_{k=1}^{n_A^W} a_k^W q^{-k}}, \tag{5}$$

$$G_H(q^{-1}, \theta_H) = \frac{\sum_{k=0}^{n_B^H} b_k^H q^{-k}}{1 + \sum_{k=1}^{n_A^H} a_k^H q^{-k}}, \tag{6}$$

where $\theta_W = [b_0^W, ..., b_{n_B^W}^W, a_1^W, ..., a_{n_A^W}^W]$ and, in a similar way, $\theta_H = [b_0^H, ..., b_{n_B^H}^H, a_1^H, ..., a_{n_A^H}^H]$. The static nonlinearity is expressed as a basis functions expansion

$$f(\theta_{NL}, v) = \sum_{k=1}^{d} \theta_{NL}^k f_k(v), \tag{7}$$

where f_k are the basis functions, θ_{NL}^k are the parameters entering linearly in f and d is the number of basis functions. In the following, it is assumed that the model structure contains the true system to identify.

2.2 Data

For the estimation of the parameter vector (4) a set of N data is assumed to be available, $\{u(t), y(t)\}_{t=1}^N$. For consistency, it is assumed that the input signal $u(t)$ is a realization of a Gaussian process and it is known exactly, while the signal $y(t)$ is the output of a true W-H system, $y_0(t)$, corrupted by stationary additive noise

$$y(t) = y_0(t) + n_y(t). \tag{8}$$

The intermediate signals $v(t)$ and $w(t)$ are not available.

2.3 General Identification Problem

A general identification problem aims to define the estimate $\hat{\theta}_N$ of the parameter vector θ. A standard prediction error method is used. It is based on minimizing the prediction error

$$\varepsilon(t, \theta) = y(t) - \hat{y}(t, \theta), \tag{9}$$

where $y(t)$ is the measured output and the prediction $\hat{y}(t, \theta)$, from (3), is given by

$$\hat{y}(t) = G_H(q^{-1}, \theta_H) f(\theta_{NL}, G_W(q^{-1}, \theta_W)u(t)). \tag{10}$$

Then the estimate $\hat{\theta}_N$ can be found by

$$\hat{\theta}_N = \text{argmin}_\theta V_N(\theta), \tag{11}$$

where $V_N(\theta)$ is the criterion of fit defined by

$$V_N(\theta) = \frac{1}{N} \sum_{t=1}^{N} \varepsilon^2(t, \theta). \tag{12}$$

Since the predictor is not a linear regression, the computation of $\hat{\theta}_N$ must be done using a gradient based iterative algorithm. That is, given a start value $\theta^{(0)}$, iterate

$$\theta^{(i+1)} = \theta^{(i)} - R_i \frac{dV_N(\theta)}{d\theta} \tag{13}$$

until convergence. Typically, $V_N(\theta)$ can have many minima and the initial estimate $\theta^{(0)}$ is crucial for the success of the minimization. Depending on R_i, (13) describes a wide class of well-known algorithms, like Gauss-Newton and Levenberg-Marquardt algorithms. In the example in Section 5 a Levenberg-Marquardt algorithm is used, implemented in a software package for the Mathematica platform, Sjöberg and Hjalmarsson (2009).

3. THE FRACTIONAL APPROACH FOR INITIALIZATION

3.1 The Best Linear Approximation

In general, the first step of the initialization procedure for the identification of W-H model is to compute the best linear model $G(q^{-1})$ of the plant. Under the assumptions of Gaussian input $u(t)$ and output $y(t)$ obtained by filtering $u(t)$ through a system of form (1)-(2)-(3) with linear parts being stable, single input, single output, finite order transfer functions, and the true nonlinear part f^0 being a continuous function $R \to R$, the best linear approximation (BLA) of the plant converges, asymptotically with N, to

$$kG_W^0(q^{-1})G_H^0(q^{-1}), \tag{14}$$

where G_W^0 and G_H^0 are the true linear functions and k is a constant value which depends on $u(t)$,f^0, G_W^0 and G_H^0 (Pintelon and Schoukens, 2001). In other words, the BLA captures the dynamics of the two linear parts and the nonlinear function is approximated with a constant. This result holds when $N \to \infty$ but, in practice, the number of data is limited and the BLA is only used to obtain the initial parameter estimate for the linear parts.

3.2 The Fractional Approach

The main idea is to use the BLA to initialize the Wiener and Hammerstein dynamics in a fractional way. The position of poles and zeros, coming from the BLA, is param-

eterized between G_W and G_H through a new vector of parameters $[\alpha, \beta]$, in the following way

$$\hat{G}_W(q^{-1}, \alpha, \beta) = \frac{\prod_{i=1}^{n_B}(1 - z_i^{BLA}q^{-1})^{\beta_i}}{\prod_{i=1}^{n_A}(1 - p_i^{BLA}q^{-1})^{\alpha_i}}, \quad (15)$$

$$\hat{G}_H(q^{-1}, \alpha, \beta) = \frac{\prod_{i=1}^{n_B}(1 - z_i^{BLA}q^{-1})^{1-\beta_i}}{\prod_{i=1}^{n_A}(1 - p_i^{BLA}q^{-1})^{1-\alpha_i}}, \quad (16)$$

where z_i^{BLA} and p_i^{BLA} are, respectively, zero and pole of the BLA. The indeces n_A and n_B denote the dimension of α and β vectors: for each real pole (zero) or pair of complex poles (zeros) only one α_i (β_i) is introduced, in order to keep the complex pairs together. In this way, $\alpha_i = 1$ ($\beta_i = 1$) locates the corresponding pole (zero) at G_W, $\alpha_i = 0$ ($\beta_i = 0$) locates the corresponding pole (zero) at G_H. Therefore, by estimating α and β, the best position is identified. The optimization problem for the estimation of α and β can be written as

$$\min_{\alpha,\beta,\theta_{NL}} V_N(\theta), \quad (17)$$

where the criterion of fit $V_N(\theta)$ is computed using the error between the system output $y(t)$ and the predictor

$$\hat{y}(t, \theta) = \hat{G}_W(q^{-1}, \alpha, \beta) f(\theta_{NL}, \hat{G}_H(q^{-1}, \alpha, \beta)u(t)). \quad (18)$$

Note that the parameters of the nonlinearity θ_{NL} are estimating at this step as well, $\theta = [\theta_{NL}, \alpha, \beta]$.

In Vanbeylen (2014), this fractional approach has been introduced and the optimization problem (17) has been solved in the frequency domain. Aim of the next sections is to present a time-domain solution for (17).

3.3 The Time-Domain Formulation

The main problem that arises in the time-domain is to find a proper formulation of the predictor (18), since, in general, α and β can be different from 0 or 1 and, thus, the time-delay operator q may have fractional exponents (see Equations (15) and (16)). A possible solution to the problem is to expand the two transfer functions G_W and G_H using the Taylor series, with respect to the delay operator q^{-1}. For example, the expansion around 0 of a transfer function with only one pole is

$$\begin{aligned}\frac{1}{(1 - pq^{-1})^{\alpha}} \approx & 1 + p\alpha q^{-1} + p^2\alpha(\alpha + 1)q^{-2} + \\ & + p^3\alpha(\alpha + 1)(\alpha + 2)q^{-3} + \cdots + \\ & + p^n\alpha(\alpha + 1)\cdots(\alpha + n - 1)q^{-n},\end{aligned} \quad (19)$$

where α is a scalar and n is the expansion order. The choice of n is crucial in this context. In order to retrieve a good approximation of the transfer functions, the expansion order needs to be increased as the model order increases. This leads, in general, to expansions containing huge algebraic expressions in α and β, see (19), that increase the computational time of the optimization procedure. Thus, the idea is to re-define the expansion order as the sum of two new indeces, $n = n_1 + n_2$, where:

- n_1 defines the number of terms in the expansion which depend on α and β;
- $n_2 = n - n_1$ is the number of remaining terms in the expansion in order to ensure a "good" approximation. These terms can be parameterized using dummy parameters d_i (not depending on α, β). "Good" is here defined as the measure of the comparison between the prediction error of the BLA and the product of the two expansions.

Using these two new indeces, the expansion of the example (19) becomes

$$\begin{aligned}\frac{1}{(1 - pq^{-1})^{\alpha}} \approx & 1 + p\alpha q^{-1} + p^2\alpha(\alpha + 1)q^{-2} + \cdots + \\ & + p^{n_1}\alpha(\alpha + 1)\cdots(\alpha + n_1 - 1)q^{-n_1} + \\ & + d_1 q^{-(n_1+1)} + \cdots + d_{n_2}q^{-n},\end{aligned} \quad (20)$$

that can be rewritten as

$$\begin{aligned}\frac{1}{(1 - pq^{-1})^{\alpha}} \approx & 1 + A_1(\alpha, p)q^{-1} + A_2(\alpha, p)q^{-2} \\ & + \cdots + A_{n_1}(\alpha, p)q^{-n_1} + \\ & + d_1 q^{-(n_1+1)} + \cdots + d_{n_2}q^{-n},\end{aligned} \quad (21)$$

where $A_1(\alpha, p), ..., A_{n_1}(\alpha, p)$ are the coefficients of the expansion depending on the known dynamics and the new parameter α. Therefore, by using the expansions, the predictor (18) becomes

$$\begin{aligned}&\hat{y}(t, \theta) = \\ &= \tilde{G}_W(q^{-1}, \alpha, \beta, d_W) f(\theta_{NL}, \tilde{G}_H(q^{-1}, \alpha, \beta, d_H)u(t)),\end{aligned} \quad (22)$$

where $\tilde{G}_W$ and $\tilde{G}_H$ represent the series expansions of $\hat{G}_W$ and $\hat{G}_H$ that, in the general case, can be expressed as

$$\begin{aligned}\tilde{G}_W(q^{-1}, \alpha, \beta) = & 1 + A_1(\alpha, \beta)q^{-1} + \cdots + A_{n_1}(\alpha, \beta)q^{-n_1} + \\ & + d_{W_1} q^{-(n_1+1)} + \cdots + d_{W_{n_2}}q^{-n},\end{aligned} \quad (23)$$

$$\begin{aligned}\tilde{G}_H(q^{-1}, \alpha, \beta) = & 1 + B_1(\alpha, \beta)q^{-1} + \cdots + B_{n_1}(\alpha, \beta)q^{-n_1} + \\ & + d_{H_1} q^{-(n_1+1)} + \cdots + d_{H_{n_2}}q^{-n}.\end{aligned} \quad (24)$$

In (23) and (24), $A_1(\alpha, \beta)$, ..., $A_{n_1}(\alpha, \beta)$, $B_1(\alpha, \beta)$, ..., $B_{n_1}(\alpha, \beta)$ are the coefficients of the Taylor series depending on the known system dynamics and on the α, β parameters, while $d_W = [d_{W_1}, ..., d_{W_{n_2}}]$, $d_H = [d_{H_1}, ..., d_{H_{n_2}}]$ are dummy coefficients. One could notice that (23) and (24) can be seen as linear regressions in $A_i(\alpha, \beta)$, $B_i(\alpha, \beta)$. With this new parameterization, the number of algebraic expressions in α, β, defined by n_1, can be kept low, without affecting the goodness of the approximation. On the other hand, n_1 should be big enough in order to have, in the predictor, a sufficient number of coefficients that allows the computation of α and β. This aspect is discussed in the next sections and, in particular, for a test example, uniqueness in α and β, once the Taylor coefficients are known, is shown.

3.4 Uniqueness in the Fractional Exponents

Assuming that the expansion order n has been chosen to ensure a good approximation of G_W and G_H, it is possible to show that values for α and β can be analytically computed. To do so, a small example is presented. It consists of a Wiener-Hammerstein system with one pole and one zero in the first linear part (a_1 and b_1 with $a_1 \neq b_1$), one pole in the second linear part (a_2), and a second degree polynomial nonlinearity in the middle:

$$
\begin{aligned}
v(t) &= \frac{(1-b_1q^{-1})}{(1-a_1q^{-1})}u(t),\\
w(t) &= f(\theta_{NL}, v(t)) = \theta_1 v(t) + \theta_2 v^2(t),\\
y(t) &= \frac{1}{1-a_2q^{-1}}w(t).
\end{aligned} \tag{25}
$$

In order to build the predictor, the first step is to compute the BLA of the whole system. Let's assume, in this example, that the BLA is known and exactly identical to the series of the two linear functions

$$
G_{BLA}(q^{-1}) = \frac{(1-b_1q^{-1})}{(1-a_1q^{-1})(1-a_2q^{-1})}. \tag{26}
$$

Two poles and one zero mean that $\alpha = [\alpha_1, \alpha_2]^T$ is a vector in R^2 and $\beta = \beta_1$ is scalar. Therefore, (15) and (16) become

$$
\hat{G}_W(q^{-1}, \alpha, \beta) = \frac{(1-b_1q^{-1})^{\beta_1}}{(1-a_1q^{-1})^{\alpha_1}(1-a_2q^{-1})^{\alpha_2}}, \tag{27}
$$

$$
\hat{G}_H(q^{-1}, \alpha, \beta) = \frac{(1-b_1q^{-1})^{1-\beta_1}}{(1-a_1q^{-1})^{1-\alpha_1}(1-a_2q^{-1})^{1-\alpha_2}}. \tag{28}
$$

One could notice that $\hat{G}_H$ is the same function as $\hat{G}_W$, but evaluated in $(1-\alpha)$ and $(1-\beta)$ instead of α, β. This property is preserved in the expansion. In fact, the generic $A_i(\alpha,\beta)$, $B_i(\alpha,\beta)$ coefficients of the expansions (23), (24) are defined as

$$
A_i(\alpha,\beta) = \frac{\hat{G}_W^{(m)}(0,\alpha,\beta)}{m!}, \quad B_i(\alpha,\beta) = A_i(1-\alpha, 1-\beta), \tag{29}
$$

where $\hat{G}_W^{(m)}(0,\alpha,\beta)$ denotes the m-th derivative of $\hat{G}_W$ with respect to the delay operator, evaluated in 0. From (23), (24) and (29) it follows that, if information about α, β can be retrieved considering $\tilde{G}_W$, the same information is retrieved for $(1-\alpha)$, $(1-\beta)$, considering $\tilde{G}_H$. Thus, it is sufficient to consider the $A_i(\alpha,\beta)$ terms, with $i = 1, ..., n_1$, in order to compute α, β.

Since the index n_1 describes how many $A_i(\alpha,\beta)$ are present in the expansion, it is reasonable to choose, in a first attempt, $n_1 = 3$ in order to have a number of coefficients equal to the number of parameters to compute $(\alpha_1, \alpha_2, \beta_1)$. With $n_1 = 3$, the predictor for system (25) can be computed as

$$
\begin{aligned}
\hat{y}(t) =& \hat{w}(t) + B_1\hat{w}(t-1) + B_2\hat{w}(t-2) +\\
&+ B_3\hat{w}(t-3),
\end{aligned} \tag{30}
$$

where $\hat{w}(t)$ is the output of the nonlinearity and can be expressed as

$$
\begin{aligned}
\hat{w}(t) =& \theta_1 u(t) + \theta_2 u^2(t) + A_1\theta_1 u(t-1) +\\
&+ 2A_1\theta_2 u(t)u(t-1) + A_1^2\theta_2 u^2(t-1) +\\
&+ A_2\theta_1 u(t-2) + 2A_2\theta_2 u(t)u(t-2) +\\
&+ 2A_1A_2\theta_2 u(t-1)u(t-2) + A_2^2\theta_2 u^2(t-2) +\\
&+ A_3\theta_1 u(t-3) + 2A_3\theta_2 u(t)u(t-3) +\\
&+ 2A_1A_3\theta_2 u(t-1)u(t-3) +\\
&+ 2A_2A_3\theta_2 u(t-2)u(t-3) + A_3^2\theta_2 u^2(t-3).
\end{aligned} \tag{31}
$$

For simplicity, in (30) and (31) the dummy parameters have been omitted. This does not affect the analysis since these parameters do not depend on α, β. The nonlinear function is linearly parameterized in θ_{NL} and the Taylor expansions of the fractional transfer functions can be seen as a linear regression where the coefficients are algebraic expressions in the unknowns α and β. Therefore it is possible to re-parameterize (30) as a linear regression of Taylor coefficients and parameters of the static nonlinearity.

For the test example, it can be shown that, from the linear regression parameters, values for the Taylor coefficients A_1, A_2, A_3, B_1, B_2, B_3 can be identified. The next step is to show that, given the Taylor coefficients, the solution in α and β is unique.

Let's denote with $\bar{A}_1$, $\bar{A}_2$, $\bar{A}_3$ the computed values for $A_1(\alpha,\beta)$, $A_2(\alpha,\beta)$, $A_3(\alpha,\beta)$. It is possible to build the following system of equations

$$
\begin{aligned}
\bar{A}_1 &= A_1(\alpha,\beta)\\
\bar{A}_2 &= A_2(\alpha,\beta),\\
\bar{A}_3 &= A_3(\alpha,\beta)
\end{aligned} \tag{32}
$$

This system can be solved analytically. Replacing the explicit expression for $A_1(\alpha,\beta), A_2(\alpha,\beta), A_3(\alpha,\beta)$, system (32) becomes

$$
\begin{aligned}
\bar{A}_1 =& a_1\alpha_1 + a_2\alpha_2 - b_1\beta_1\\
\bar{A}_2 =& \frac{1}{2}(a_1^2\alpha_1(\alpha_1+1) + 2a_1a_2\alpha_1\alpha_2 +\\
&+ a_2^2\alpha_2(\alpha_2+1) - 2(a_1\alpha_1 + a_2\alpha_2)b_1\beta_1 +\\
&+ b_1^2(\beta_1-1)\beta_1)\\
\bar{A}_3 =& \frac{1}{6}(a_1^3\alpha_1(\alpha_1+1)(\alpha_1+2) +\\
&+ 3a_1^2a_2\alpha_1(\alpha_1+1)\alpha_2 +\\
&+ 3a_1a_2^2\alpha_1\alpha_2(\alpha_2+1) +\\
&+ a_2^3\alpha_2(\alpha_2+1)(\alpha_2+2) +\\
&- 3(a_1^2\alpha_1(\alpha_1+1) + 2a_1a_2\alpha_1\alpha_2 +\\
&+ a_2^2\alpha_2(\alpha_2+1))b_1\beta_1 +\\
&+ 3(a_1\alpha_1 + a_2\alpha_2)b_1^2(\beta_1-1)\beta_1 +\\
&- b_1^3(\beta_1-2)(\beta_1-1)\beta_1)
\end{aligned} \tag{33}
$$

The system (33) consists of three polynomial equations in $\alpha_1, \alpha_2, \beta_1$. The first equation of system (33) is linear. Thus, it is possible to compute, for example, α_1 in function of α_2 and β_1

$$
\alpha_1 = \frac{\bar{A}_1 - a_2\alpha_2 + b_1\beta_1}{a_1}. \tag{34}
$$

Replacing (34) in the second equation of (33), the quadratic terms in α_2, β_1 are simplified and the equation becomes

$$
\bar{A}_2 = \frac{1}{2}(a_2^2 - a_1a_2)\alpha_2 + \frac{1}{2}(\bar{A}_1^2 + a_1\bar{A}_1 + (a_1b_1 - b_1^2)\beta_1), \tag{35}
$$

linear in α_2, β_1. Thus, it is possible to compute

$$
\alpha_2 = \frac{\bar{A}_1^2 + a_1\bar{A}_1 - 2\bar{A}_2 + (a_1b_1 - b_1^2)\beta_1}{(a_1 - a_2)a_2}. \tag{36}
$$

Replacing (34) and (36) in the third equation of (33), we get a final linear equation in β_1

$$
\begin{aligned}
\bar{A}_3 =& \frac{1}{3}(a_1b_1^2 + a_2b_1^2 - b_1^3 - a_1a_2b_1)\beta_1 +\\
&+ \frac{1}{3}(2a_1\bar{A}_2 + 2a_2\bar{A}_2 + 3\bar{A}_1\bar{A}_2 +\\
&- a_1a_2\bar{A}_1 - a_1\bar{A}_1^2 - a_2\bar{A}_1^2 - \bar{A}_1^3).
\end{aligned} \tag{37}
$$

Unique solution in $\alpha_1, \alpha_2, \beta_1$ is retrieved and it can be expressed as

$$\alpha_1 = \frac{K_1 + K_2 b_1 + (K_3 b_1 + K_4) a_2 + K_5}{a_1(a_1 - a_2)(a_1 - b_1)}$$
$$\alpha_2 = -\frac{K_1 + K_2 b_1 + (K_3 b_1 + K_4) a_1 + K_5}{a_2(a_1 - a_2)(a_2 - b_1)} \quad (38)$$
$$\beta_1 = \frac{K_1 + K_2 a_2 + (K_3 a_2 + K_4) a_1 + K_5}{b_1(a_1 - b_1)(a_2 - b_1)}$$

where K_1, K_2, K_3, K_4, K_5 are constant values depending on $\bar{A}_1, \bar{A}_2, \bar{A}_3$

$$\begin{aligned} K_1 &= \bar{A}_1^3 \\ K_2 &= \bar{A}_1^2 - 2\bar{A}_2 \\ K_3 &= \bar{A}_1 \\ K_4 &= K_2 \\ K_4 &= 3\bar{A}_3 - 3\bar{A}_1\bar{A}_2 \end{aligned} \quad (39)$$

The solution (38) points out a drawback of the approach. In fact, if double poles are present or one pole simplifies with a zero, at least two of the denominators in (38) will be zero. This simple example shows that it is possible to uniquely identify the parameter vectors α, β from a system of equations built up using the coefficients of the Taylor series. It also shows that, for uniqueness, index n_1 has to be chosen equal to the sum of the dimensions of vectors α, β to identify.

4. THE IDENTIFICATION ALGORITHM

On a test example, the previous section has provided results regarding uniqueness of the solution in the α, β parameters. This, together with the identification of the nonlinear parameters θ_{NL}, provides an initial guess $\theta^{(0)}$ for the iterative optimization. The values of α and β obtained with this approach explain how the dynamics are divided in the two linear parts. The next step of the identification procedure consists in placing the poles and zeros in the first or second linear part, according to the values of α and β, and re-fitting both dynamics and nonlinear parameters. Motivation for that is the fact that the BLA, as already explained in Section 3.1, is only an approximation of the true dynamics of the system, since the number of data is limited. In the following, a summary of the main steps of the identification procedure is presented.

(1) Estimate the BLA of the plant as a rational form.
(2) Choose indeces n and n_1, and expand the two fractional functions (15) and (16).
(3) Initialize the two expansions with $\alpha_i = \frac{1}{2}$, $i = 1, ..., n_a$ an $\beta_j = \frac{1}{2}$, $j = 1, ..., n_b$.
(4) Minimize the cost function (12) and determine the values for α, β, where α, β appear as unknown in the Taylor coefficients. The nonlinear parameters θ_{NL} are estimated at this step as well. This can be implemented through the standard Levenberg-Marquardt algorithm (Marquardt, 1963).
(5) Place poles and zeros according to the computed values of α_i and β_j and re-parameterized them as well as the nonlinearity.
(6) Re-fit poles and zeros and nonlinear parameters.

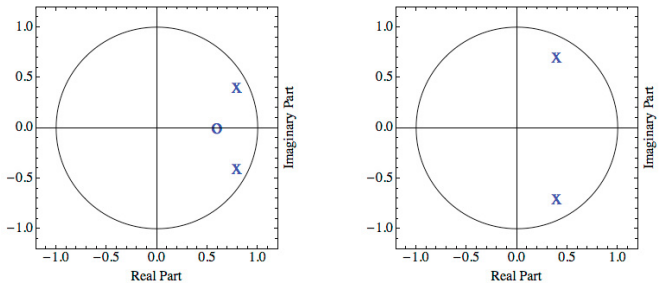

Fig. 2. Poles (x) and zeros (o) of the two linear parts of the true system.

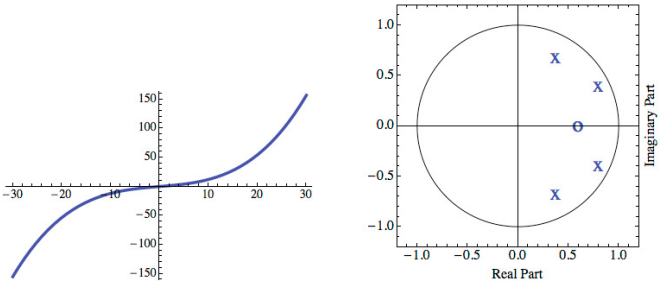

Fig. 3. a) True nonlinearity. b) Poles (x) and zeros (o) of the BLA.

5. RESULTS

5.1 Simulation Example

In this section the proposed algorithm is tested on data generated by W-H model. A white Gaussian signal with standard deviation 5 is used as input signal to a system with the W-H structure. Figure 2 depicts the poles and zeros of the two linear parts and Figure 3a) shows the true nonlinearity. With this true system, 3000 data samples were generated. The output is corrupted with white Gaussian noise with standard deviation 0.1. The steps of the identification algorithm are described in the following.

(1) The BLA is estimated. For that, standard linear system idendentification algorithms are used. The poles and zeros of the BLA are shown in Figure 3b).
(2) Poles and zeros from the BLA are used to initialize the fractional functions (15) and (16). Since the BLA consists of two pairs of complex conjugate poles and one real zero, two α and one β have to be introduced. The expansion of (15) and (16) can be performed. The index n_1 is chosen to be equal to 3 (two α and one β), while $n = 14$ gives a root mean square error of 18.66, close to the one of the BLA (18.52).
(3) Parameters α_1, α_2, β_1 are initialized with 0.5.
(4) The cost fuction (12) is minimized, obtaining values for α, β ($\alpha_1 = -0.11424$, $\alpha_2 = 1.05687$, $\beta_1 = 1.16543$) and θ_{NL}. Values of α, β are classified to one or zero: $\alpha_1^* = 0$, $\alpha_2^* = 1$, $\beta_1^* = 1$.
(5) According to α_1^*, α_2^*, β_1^*, the position of poles and zeros is defined and they can be re-parameterized for the final optimization step.

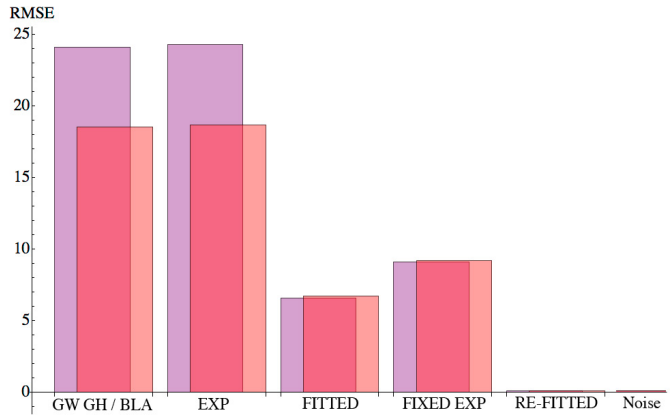

Fig. 4. Root Mean Square Errors of the identification procedure when the BLA (front) and the product $G_W\,G_H$ (back) is used to initialize the model structure. From the left: 1) RMSE of the BLA and $G_W\,G_H$. The BLA shows a lower error, since it describes the best linear model of the system; 2) RMSE of the product of the expansions; 3) RMSE of the W-H model after the fitting of α, β and θ_{NL}. In this case the model initialized with product $G_W\,G_H$ performs better, since the nonlinearity is now fitted and the position of poles and zeros, for this model, is the true one; 4) RMSE of the W-H model where α and β are classified to {0,1}; 5) RMSE of the W-H model where poles and zeros are re-fitted together with the nonlinear parameters; 6) Noise level.

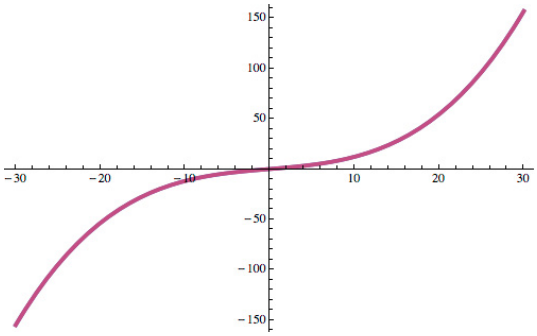

Fig. 5. True and identified nonlinearity. The two curves are overlapping.

(6) Poles, zeros and nonlinear parameters are re-fitted. The result is a W-H model whose RMSE is comparable with the noise level.

In Figure 4, the prediction errors of the main steps of the identification procedure are shown. For comparison, the prediction errors of the identification procedure when the product of the true G_W and G_H is used, instead of the BLA, are shown as well. Finally, Figure 5 shows the comparison between the true nonlinearity and the identified one.

The key point of the present approach is that the position of poles and zeros from the BLA is parameterized and the best position is retrieved. In order to compare this result with the approach in Wills and Ninness (2012), the position of the two complex pairs of poles of this example has been randomly initialized between the first and the second linear parts. The result of the test is that in the 25% of the cases the poles were not able to "move" to the best position, showing that a local minimum was found.

6. CONCLUSIONS

A time-domain formulation of the fractional approach for initializing W-H models has been introduced. With this formulation the initialization problem consists of two main parts:

- using Taylor expansion, the time-domain predictor is written as a linear regression where the coefficients are algebraic expressions in the fractional exponents and in the parameters of the static nonlinearity. This result holds in general, under the assumption of linearly-parameterized nonlinearity;
- given the coefficients, values for the algebraic expressions can be identified. This allows to analytically compute the fractional exponents, meaning that the position of poles and zeros is determined. On a low order example, uniquess of the solution is shown. Generalization of this result to any model order is addressed in future research.

This approach allows to retrieve a good initial estimate which does not suffer of local minima problems. The initial estimate is then used for the final optimization in all model parameters.

REFERENCES

Ljung, L. (ed.) (1999). *System Identification (2Nd Ed.): Theory for the User*. Prentice Hall PTR, Upper Saddle River, NJ, USA.

Marquardt, D.W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *SIAM Journal on Applied Mathematics*, 11(2), 431–441.

Pintelon, R. and Schoukens, J. (2001). *System identification: a frequency domain approach*. IEEE press, New York.

Sjöberg, J. and Hjalmarsson, H. (2009). A system, signals and identification toolbox in mathematica with symbolic capabilities. *In 15th IFAC Symposium on System Identification, Saint-Malo, France*.

Sjöberg, J., Lauwers, L., and Schoukens, J. (2012). Identification of WienerHammerstein models: Two algorithms based on the best split of a linear model applied to the sysid'09 benchmark problem. *Control Engineering Practice*, 20(11), 1119 – 1125. Special Section: Wiener-Hammerstein System Identification Benchmark.

Sjöberg, J. and Schoukens, J. (2012). Initializing Wiener-Hammerstein models based on partitioning of the best linear approximation. *Automatica*, 48(2), 353–359.

Söderström, T. and Stoica, P. (eds.) (1988). *System Identification*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA.

Vanbeylen, L. (2014). A fractional approach to identify Wiener-Hammerstein systems. *Automatica*, 50(3), 903 – 909.

Wills, A. and Ninness, B. (2012). Generalised Hammerstein-Wiener system estimation and a benchmark application. *Control Engineering Practice*, 20(11), 1097 – 1108. Special Section: Wiener-Hammerstein System Identification Benchmark.